

Intro to Translation

Introduction

DNA gave us mRNA. mRNA came out of the nucleus and into the cytoplasm. Now it's time to make some proteins. Proteins are a series of amino acids tethered together. The genetic code is defined by the relationship of the nucleotides in the DNA (which are the same as the mRNA, only T's turned to U's), and the sequence of the amino acids. Each amino acid is linked to a codon. A codon is a 3-nucleotide sequence. Ribosomes (rRNAs) provide the machinery to translate the copy of the genetic code, the mRNA sequence, into a protein. rRNAs read the mRNA sequence. They pair tRNA with that mRNA sequence. The tRNA is carrying one amino acid. The rRNA ensures that the amino acids get connected, the end of the chain of amino acids already made to the new incoming amino acid, added to the ever-growing chain of amino acids. This lesson gives a high-level overview, defines terms, and lays a foundation for the subsequent lessons, which have a lot of detail.

Orientation of Amino Acids to DNA and mRNA

Like every step in gene expression, translation has some orientation issues. Go to any source and you'll see them flip the directionality on you, which makes learning really hard. I said we'd always have the coding DNA strand 5' left, 3' right. That made the mRNA strand 5' left, 3' right. Keeping with that model, the **amino acid sequence** of translation should have a 5' left, 3' right orientation.

That means that the original sequence of the coding strand, read 5' left to 3' right, just as we read English left to right, will show the amino acids in order for the protein each will code (except introns).

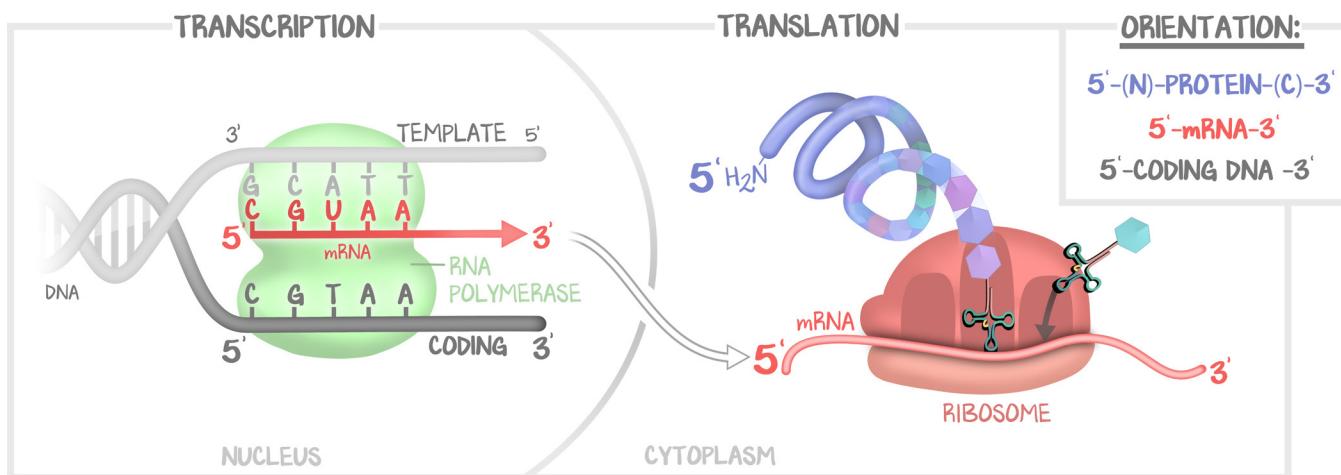


Figure 12.1: Orientation

The coding strand is kept 5' left to 3' right, which corresponds to the mRNA's 5' left to 3' right, which then correlates to the amino acid sequence's 5' left to 3' right.

Only in proteins there isn't a 5' end and a 3' end. Amino acids are bigger than that. Instead, they have an Amino end (N-terminus) and a Carboxy end (C-terminus). **5' = N-terminus, 3' = Carboxy terminus.** Ensure that the mRNA is always 5' left, 3' right; ensure that the amino acid sequence is N-term left, C-term right.

This will make anticodon reading go backwards. The only drawback to this method is if we need to translate TO an anticodon ... something we shouldn't have to do—tRNA does it for us. With lookup tables, we use the codon, the mRNA sequence, 5' to 3' ... NOT the anticodon of the tRNA. But if you get a tricksie professor (the USMLE doesn't care about your ability to read backwards), keep an eye out for directionality.

Reading Frames

When we were reading DNA and RNA, we read one nucleotide at a time. That's how it works. That's how DNA polymerase works. That's how RNA polymerase works. Well, these are small proteins working with small strands. Ribosomes, rRNA, are small proteins working with large strands of RNA to build really big proteins (relative to the strand of RNA). Which means that ribosomes are going to read **three nucleotides at a time**. This three-nucleotides-at-a-time unit is called a **codon**.

The figure is a grid representing the genetic code. The columns are labeled U, C, A, and G, representing the first base. The rows are also labeled U, C, A, and G, representing the second base. The third base is indicated by the color of the amino acid names: purple for non-stop codons and red for stop codons (UAA, UAG, UGA). The fourth column, G, represents the third base.

	U	C	A	G
1ST BASE	UUU PHE UUC UUA UUG	UCU SER UCC UCA UCG	UAU TYR UAC UAA STOP UAG STOP	UGU CYS UGC UGA STOP UGG TRP
	CUU CUC LEU CUA CUG	CCU PRO CCC CCA CCG	CAU HIS CAC CAA GLN CAG	CGU CGC ARG CGA CGG
	AUU ILE AUC AUA AUG MET	ACU THR ACC ACA ACG	AAU ASN AAC AAA LYS AAG	AGU SER AGC AGA AGG ARG
3RD BASE	GUU GUC VAL GUA GUG	GCU ALA GCC GCA GCG	GAU ASP GAC GAA GLU GAG	GGU GGC GLY GGA GGG

Figure 12.2: Codon Chart

tRNA is a cloverleaf structure that carries an amino acid on its 3' end. At the very tip of the cloverleaf is a sequence of three nucleotides called an anticodon. This anticodon is complementary and antiparallel to the mRNA it is reading. Like the template strand, the tRNA anticodon is backwards and paired.

Codes and Codons

The genetic code is **unambiguous**, which means that **one codon** codes for **one amino acid**. The same codon codes for the same one amino acid. That is true of every living organism (a profound statement), so the genetic code is universal.

The code is also said to be **degenerate**, a term that really means redundant. Since there are 4 base pairs, and every codon is 3 base pairs long, there are a total of $4 \times 4 \times 4$ permutations, which is 64 codons. There are only **20 amino acids**, which includes a start (**methionine**), and three stop codons (UAG, UAA, UGA). With 20 amino acids and 3 stops, there are many more codons to go around than there are amino acids. This means that amino acids have more than one codon. The code is unambiguous—one codon, one amino acid. The code is degenerate—one amino acid may have multiple codons. The word “degenerate” is obnoxious. It doesn’t help understand what it’s referring to. “Redundant” makes more sense. But there is more to it than simple redundancy, as we’ll see when we discuss wobble codons in #13: *Translation*. For now, know that for any one amino acid there are multiple codons, but any time

multiple codons code for an amino acid, the **first two base pairs are conserved**. The third nucleotide “doesn’t matter as much” and is referred to as a **wobble codon** because the first two dictate the amino acid, and the third confirms. The third codon does change the amino acid, and this happens a lot, but any time there is more than one codon for one amino acid, the third nucleotide is the one that varies.

Codons are read three nucleotides at a time. These codons are **nonoverlapping** with their neighbors; the RNA can only read 1 codon, then 1 codon, then 1 codon; which is 3 nucleotides, then 3 nucleotides, then 3 nucleotides. The reading of codons is also **comma-less** (which means continuous). Ribosomes don’t skip a nucleotide, don’t miss a nucleotide, and they don’t reuse any nucleotide already used.

But **where one starts shifts the frame**. Starting on nucleotide 1 will result in a new codon at 4, 7, and 10. Starting on nucleotide 2 will result in a new codon on 5, 8, and 11. This becomes relevant in mutations. Because we have a start signal AUG, and that’s the only start we have, that mRNA is only going to make one protein. That one mRNA might have a mutation, a splice, or some other thing happen that lets the DNA code make a different protein, but the mRNA that we sent out of the nucleus is only going to make **one protein**, and that’s the protein it codes for.

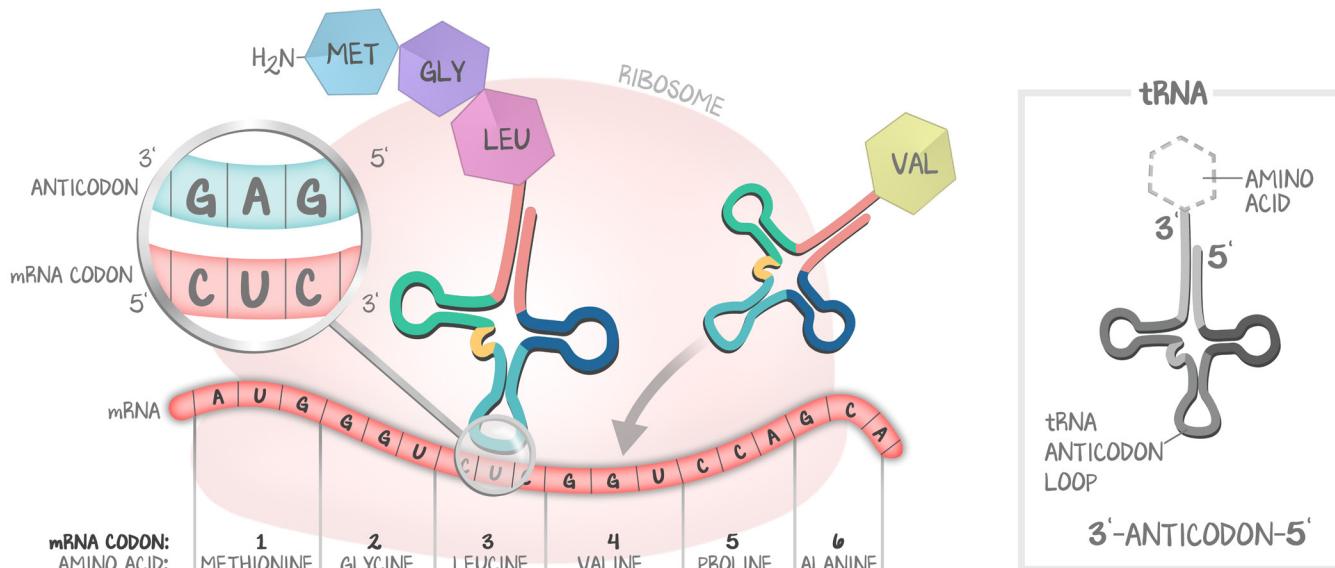


Figure 12.3: tRNA, Codon, Anticodon

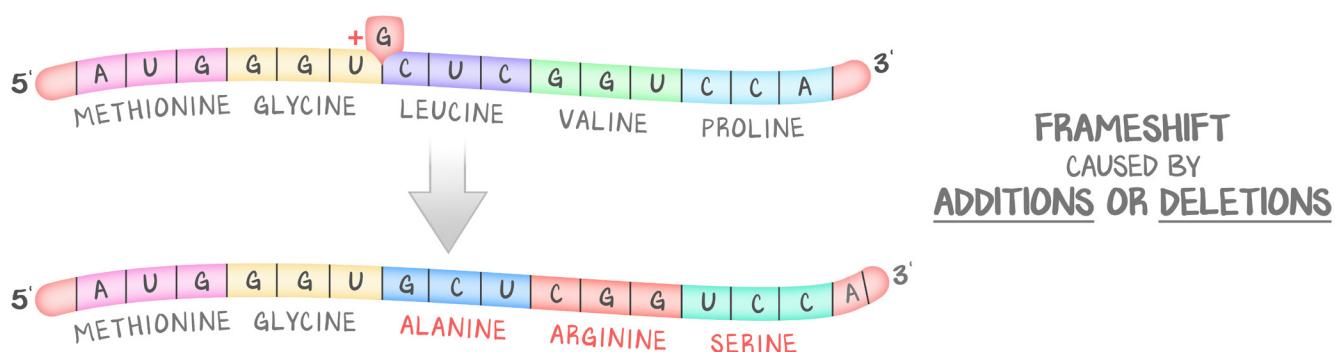


Figure 12.4: Reading Frames

The Process of Translation

rRNA forms a complete unit of two subunits around the mRNA. It knows to start when it reaches the **start codon** (**AUG** = methionine). Methionine is always the first amino acid in any sequence. Then it holds one tRNA with the elongated strand at one site, invites the incoming tRNA with the next amino acid into the other site, and catalyzes a reaction whereby the previous tRNA is booted while the entire strand of amino acid is transferred to the new tRNA, while the ribosome moves over 3 nucleotides to let another tRNA come in to the A site, now empty. This process repeats itself, 3 codons at a time, elongating the chain of amino acids until a **stop codon** is reached. See this as exactly the same thing as DNA replication or RNA transcription, except now the rRNA and tRNA are reading 3 nucleotide segments at a time instead of one. This is discussed in immense detail in #13: *Translation*.